JOHNSON GRANT
IN-61
7309
P-46

EXPERT SYSTEM VERIFICATION AND VALIDATION SURVEY

Delivery 3 - Recommendations

International Business Machines Corporation

August 31, 1990

Cooperative Agreement NCC 9-16
Research Activity No. AI.16

NASA Johnson Space Center
Information Systems Directorate
Information Technology Division

Research Institute for Computing and Information Systems
University of Houston - Clear Lake

T·E·C·H·N·I·C·A·L R·E·P·O·R·T

The RICIS Concept

The University of Houston-Clear Lake established the Research Institute for Computing and Information systems in 1986 to encourage NASA Johnson Space Center and local industry to actively support research in the computing and information sciences. As part of this endeavor, UH-Clear Lake proposed a partnership with JSC to jointly define and manage an integrated program of research in advanced data processing technology needed for JSC's main missions, including administrative, engineering and science responsibilities. JSC agreed and entered into a three-year cooperative agreement with UH-Clear Lake beginning in May, 1986, to jointly plan and execute such research through RICIS. Additionally, under Cooperative Agreement NCC 9-16, computing and educational facilities are shared by the two institutions to conduct the research.

The mission of RICIS is to conduct, coordinate and disseminate research on computing and information systems among researchers, sponsors and users from UH-Clear Lake, NASA/JSC, and other research organizations. Within UH-Clear Lake, the mission is being implemented through interdisciplinary involvement of faculty and students from each of the four schools: Business, Education, Human Sciences and Humanities, and Natural and Applied Sciences.

Other research organizations are involved via the "gateway" concept. UH-Clear Lake establishes relationships with other universities and research organizations, having common research interests, to provide additional sources of expertise to conduct needed research.

A major role of RICIS is to find the best match of sponsors, researchers and research objectives to advance knowledge in the computing and information sciences. Working jointly with NASA/JSC, RICIS advises on research needs, recommends principals for conducting the research, provides technical and administrative support to coordinate the research, and integrates technical results into the cooperative goals of UH-Clear Lake and NASA/JSC.

EXPERT SYSTEM VERIFICATION AND VALIDATION SURVEY

Delivery 3 - Recommendations

Preface

This research was conducted under auspices of the Research Institute for Computing and Information Systems by the International Business Machines Corporation. Dr. Terry Feagin and Dr. T. F. Leibfried served as RICIS research representatives.

Funding has been provided by Information Technology Division, Information Systems Directorate, NASA/JSC through Cooperative Agreement NCC 9-16 between NASA Johnson Space Center and the University of Houston-Clear Lake. The NASA technical monitor for this activity was Chris Culbert, of the Software Technology Branch, Information Technology Division, Information Technology Directorate, NASA/JSC.

Expert System Verification and Validation Survey
RICIS Contract #069
Delivery 3 - Recommendations

August 31, 1990

IBM

3700 Bay Area Blvd.
Houston, Tx. 77058

Survey Results

Preface

This document constitutes the third delivery, "Recommendations," of the four deliveries scheduled for RICIS contract 069, "Verification and Validation of Expert Systems Study." The remaining delivery is the "Final Report," due on September 14, 1990.

This delivery consists of an update to the second delivery, "Survey Results" and is reported via a new section in this document, "Recommendations" on page 20. This delivery also includes an updated "Summary of Results" section which reflects all questionnaires received as of August 29, 1990.

The final delivery will consist of an update to this document. The "Final Report" will report survey data gathered late in the contract period via updates to the "Summary of Results," and may also include minor alterations to "Recommendations" based on this new data.

Contents

Background

The purpose of this task is to determine the state-of-the-practice in Verification and Validation (V&V) of Expert Systems (ESs) on current NASA and Industry applications. This is the first task of a series which has the ultimate purpose of ensuring that adequate ES V&V tools and techniques are available for Space Station Knowledge Based Systems development.

The strategy for determining the state-of-the-practice is to check how well each of the known ES V&V issues are being addressed and to what extent they have impacted the development of Expert Systems.

Note: This task does not attempt to prove or disprove whether Verification and Validation can or should be performed on Expert Systems. It is accepted that Verification and Validation should be applied to all software systems, including Expert Systems.

Survey Rationale

It is widely claimed that Expert Systems have been not been subject to the same level of Verification and Validation as traditionally developed software. Some people feel that this lack of V&V continues because of a "vicious circle," where nobody requires expert system V&V, so nobody does it. Consequently, since nobody knows how to do it, nobody requires it. There are two major reasons why the V&V process has not been documented: lack of a single life-cycle model, and technical differences between traditional software and expert systems.

Most expert system development life-cycles rely on iterative prototypes to develop the system behavior. This approach does not lead to methodical capture and documentation of the expected system behavior. Documented expectations, traditionally captured in a requirements document, are essential in the V&V process: you can't do testing if you don't know what to test for! One goal of this survey is to understand how the expected behavior of current expert systems is communicated and evaluated, even if a formal requirements document was not developed.

Expert Systems are typically composed of three parts: the knowledge base (KB), the inference engine, and the interface code between the inference engine and the peripheral devices (terminals, sensors, effectors, users, etc.). The inference engine and interface code are simply traditional software and should currently be V&Ved by accepted practices. This survey will help determine if these parts are V&Ved or whether, since they are part of an expert system, V&V is overlooked.

The knowledge base is the only part of the Expert System that raises new and unique issues. A set of of the *possible* issues are:

Issues primarily due to use of nonprocedural languages

- Understandability and readability to support inspections
- Testing coverage
- Standard validation tests for inference engines
- Real-time performance analysis

Issues due to heuristic knowledge (difficulty in organizing)

- Knowledge validation
- Modularity/Design

Issues primarily due to solving new complex problems

- Requirements
- Certification

Other issues

- Uncertainty Analysis
- Inheritance Process Test and Analysis
- Configuration Management

One of the purposes of this survey is to find out if these identified *possible* issues actually cause problems in practice, and if so, how the issues are being handled.

Purpose of the Questionnaires

Some of the information for this survey can be captured fairly easily and is accomplished through use of a questionnaire. The information captured this way includes:

- Application information - What kind of problem does the system address?, What are the performance goals?
- Expertise information - What was the relationship between the developers and expert(s)?, What is the performance level of the expert?
- Development information - How was the system developed?, How big is the system?
- Evaluation information - How was the system evaluated?
- Performance information - How important is good performance?, How well is the ES performing?

Purpose of the Interviews

The questionnaire answers lead to an additional set of questions involving the V&V issues described earlier. The additional questions are greatly affected by the answers provided in top questionnaire, so it would be more efficient to derive the information through direct interviews than to generate a large number of secondary questionnaires. The interviews attempt to uncover:

- the real issues involved in ES V&V (in comparison with the known *possible* issues outlined above).
- what is being done currently to address V&V (inspections, path testing, testing by the expert).
- what makes users trust the ESs, if the ESs are indeed trusted.
- what problems, unique to ESs, were encountered and possibly addressed during development and test.

The interviews are also required because we expect that some people will not fill out the questionnaires.

Survey Administration

This survey was designed so that the majority of the information would be gained from direct interviews with people involved in ES projects. Several people from each project, including developers, users, and managers, were interviewed to get a realistic view of the projects.

Several other activities were undertaken, both before and after the interview activity, to ensure that the results of the survey reflected the actual "state-of-the-practice". These activities included:

Identifying candidate ES projects
A list of projects to be contacted was created. The list included projects at NASA and IBM as well as projects from fields outside of the space industry.

Developing survey questionnaire(s)
To improve the chances of getting meaningful data from the questionnaire activity, separate questionnaires were developed for developers and users. Each questionnaire includes a question to indicate if the answers are from a manager or non-manager. Questionnaires are listed in Appendix A, "Expert Systems Evaluation Questionnaire (Developer)" on page 25 and Appendix B, "Expert Systems Evaluation Questionnaire (User)" on page 33.

Evaluating returned questionnaires
Each questionnaire was evaluated to determine if project interviews would uncover more information. If a project was to be interviewed, the questionnaire results provided guidance on which topics would be the most useful to explore.

Summarizing interview/questionnaire results
The summarized results of the questionnaire/interview activities are presented in section "Summary of Results" on page 7.

Recommendations
Recommendations for further action, based on the information in "Summary of Results" on page 7 will be provided as the next delivery.

Survey Questionnaires

Different versions of the questionnaire were developed for developers and users of the expert system. In addition, responses were expected to be different between managers and non-managers, so an indication is included on each questionnaire.

Information Gathered

Several types of information are captured by the questionnaire. Each question in the questionnaire addresses at least one of the previous types of information. For each type of information, the subtopics and questions which provide information are listed. The question numbers are noted as (development question, user question). Questions not available on a questionnaire are indicated by a "-".

General Information

Describes the general properties of the expert system, including the name (1, 41), a short description (4, 44), field of the problem (5, 45), and the type of problem to be solved (6, 46). Also captured are whether the survey taker was a manager (2, 42).

Performance Criteria

A major expertise issue is performance (probability that the results given are correct); specifically performance of the experts (10, 49), expected performance of the system (11, 50), and actual performance of the system (12, 51). Related to the performance issue is the amount of the problem space that the ES is expected to cover (8, 47), and that it actually covers (9, 48).

Requirements Definition

Requirements definition information includes how the requirements are documented (13, -), the difficulty in determining the requirements (14, -), and the availability of the expert(s) to resolve requirements issues during development (17, -). Influencing the performance issue is the number of experts (15, -), and whether the experts agree on the results obtained from the system (16, 61). It may also be useful to know if the expert (-, 52) and/or the developer(s) (18, 53) are part of the user organization.

Development Information

Development information that we are concerned with includes the development life-cycle used (19, -), and what languages and tools were used to develop the system (20, -). The size of the system (22, -), the total effort required for development, (29, -), and the effort required to develop the different parts of the ES (21, -) indicate the difficulty of the development effort. The sensitivity of the system (24, -) will influence the difficulty of future maintenance activities.

V&V Activities Performed

The major information to be captured during this task is the current state-of-the-practice for V&V of ESs, including the kinds of V&V being attempted, both during (28, -) and after (33, 60) development, and how much of the development effort was spent on V&V (30, -). Detailed information is also gathered for V&V activities for Knowledge Structures (25, -), the Inference Engine (26, -), and the Interface Code (27, -).

Information about the difficulty of the V&V effort (35, 62), whether a separate group performed V&V, (31, -) and how much effort was expended on the independent V&V (32, 59), is also gathered.

Whether the system is operational or prototype (3, 43), and the criticality of the system (37, 55) have an affect on the amount of V&V activities performed.

V&V Issues Encountered

If the state-of-the-practice is to be improved, the major issues that need to be addressed must be identified. One question (36, 63). directly asks whether each the known issues was actually encountered. Additional questions find out more information about specific issues, including the existence of certainty factors (7, -), whether configuration management was performed (34, -), and the difficulty of implementing the expertise through the Knowledge Structures (23, -). User acceptance is the ultimate test of the V&V activities. The comparison between expected system use (39, 57) and actual system use (40, 58), the perceived reliability of the system (38, 56), and why the user is convinced that the system produces correct results (-, 54) are all indicators of user acceptance.

Human Factors

The questionnaires were designed to capture as much accurate information as possible. In an effort to accomplish this, the following human factors issues were taken into account:

Questions should be understandable

Questions should have as few "technical" terms as possible to avoid confusion due to local usage. For questions that must have technical content, be sure to provide sufficient explanation.

Choices worded positively

Negatively worded choices may not get selected because the responder may feel there is something wrong with it.

Meaningful questions

The responder should feel that there is some purpose to the question.

Make use of fill-in-the-blank questions

The responder should not have to fill in long responses. Some questions can not have all possible responses enumerated, so the the user should be able to specify his own choice.

Summary of Results

The survey results are summarized in the following sections. The results are organized according to the type of information, as organized in "Information Gathered" on page 5. The numbers corresponding to the developer and user questionnaires, respectively, are given for each question. If the question is not in one of the questionnaires, the position is filled with a '-' (for example, if a question was number 10 in the developers questionnaire and not in the user questionnaire, the question numbers would be given as: 10, -). The total number of responses is also given for each question. The number of times each choice was selected is given to the left of the choice.

The following is a short summary of each type of information gathered.

Note: The number of respondents has roughly doubled (from 19 to 35) since the "Survey Results" were reported on August 15th. With few exceptions, the distributions of the responses has not changed significantly. These exceptions are noted in the following summary where applicable.

Note: Not included in this summary is the information gathered for internal IBM expert systems, which currently has eighteen participants.

General Information

Most of the respondents were involved with Expert Systems which perform Diagnosis (82%) in the Aerospace field (74%). The survey respondents were predominantly involved with development (89%).

Performance Criteria

The levels of performance and problem space coverage that were expected and realized were lower than expected. The expected performance of the systems was nearly as high as the expert performance, but the actual performance was generally lower. The expected problem space coverage was not especially high; however, actual coverage was considerably less.

Requirements Definition

Of thirty respondents, twenty-four indicated that expert consultation was a basis for determining the behavior of the system. More revealing is that sixteen indicated consultation as the primary basis, while only sixteen indicated that there were any documented requirements. Fourteen respondents indicated that prototypes or similar tools were used for requirements.

Determining requirements had average difficulty. Availability of experts and agreement among experts were not problems.

Note: While expert consultation was still important, a much higher number of respondents indicated that other requirements sources were available. Also, the number of respondents which indicated that the experts were NOT the primary source for requirements increased from 13% to 20%.

Development Information

The most frequent (40%) Life-Cycle model used is the Cyclic Model (repetition of Requirements, Design, Rule Generation, and Prototyping until done); however, 27% of the respondents stated that no model was followed. Most development was done with an Expert System shell

(CLIPS and others), and the predominant Interface Code was C and LISP. Applications were reasonably large and required an average of 42 person/months to develop. Developed systems were not reported to be particularly sensitive to change.

Note: The number of respondents indicating that no life-cycle model was followed increased from 19% to 27%. This is surprising since the percentage of operational systems (as noted below) also *increased* from 37% to 46%.

V&V Activities Performed

Most V&V activities relied on comparison with expected results and expert checking. Typically, 19% of the development effort was spent on V&V. The difficulty of the V&V effort was reported to be medium.

In most cases, there was not a separate group to perform V&V. When reported, the V&V effort expended varied widely between developers (1.7 person/months) and users (16 person/months). Fifty-three percent of the respondents indicated that the ES was a prototype system.

Note: In addition to the increase in operational systems from 37% to 47%, much less reliance on experts to perform testing was reported, and the V&V effort was reportedly harder.

V&V Issues Encountered

The known issues most often cited as problems were: knowledge validation (66%), test coverage determination (59%), and problem complexity (50%). The least cited problem was analysis of certainty factors (only two respondents indicated that certainty factors were used). Every known issue was cited by at least one respondent.

Configuration management practices are reported to be an issue for many participants, regardless of whether the system was operational or a prototype. The expected system use varied widely (3-2000), while actual system use was relatively good (less than half of the respondents provided information, suggesting that actual use was much lower than reported). System reliability, and expertise implementation difficulty were about average.

Note: The incidence of several issues changed significantly, probably due to the emphasis on more operational systems:

- Modularity/Design of knowledge structures is much more significant, with 34% reporting problems, versus 19% earlier.
- Configuration Management is more of a concern, appearing on 20% of the questionnaires, versus 6% earlier.
- The overall difficulty of implementing the expertise is slightly lower when the additional data is considered.

General Information

The questions for the name of the ES, and the short description are not reported.

Field of the Problem

Question Numbers: 5, 45
Total Responses: 35

What field does the problem belong to?

26 Aerospace
_2 Financial
__ Information Systems
__ Hardware
__ Manufacturing
__ Marketing
__ Medical
__ Personnel
_1 Research
__ Service
_1 Software
_5 Other

Type of Problem Solved

Question Numbers: 6, 46
Total Responses: 34

Which of the following items best describes the kind of problem the Expert System addresses? Please indicate primary purpose with a '*' and check all other applicable purposes (if any).

Note: The number of times the choice was selected as primary purpose is given in parentheses after the number of times the choice was selected.

_5 (_4) Design - Configuring objects under constraints
_5 (__) Repair - Executing plans to administer prescribed remedies
_8 (_4) Control - Governing overall system behavior
_9 (_1) Planning - Designing actions
28 (14) Diagnosis - Inferring system malfunctions from observables
_6 (__) Debugging - Prescribing remedies for malfunctions
13 (__) Prediction - Inferring likely consequences of given situations
17 (_2) Monitoring - Comparing observations to expected outcomes
_5 (__) Instruction - Diagnosing, debugging, and repairing behavior
10 (_2) Interpretation - Inferring situation descriptions from sensor data
_2 (_1) Classification - Categorizing objects by properties data

Role on Project

Question Numbers: 2, 42
Total Responses: 35

Were you a developer of the Expert System the manager of the, development organization, a user of the Expert System, or the manager of a department which uses the Expert System?

15 Developer of Expert System
_6 Manager of Expert System development organization
10 Other Development
_4 User of the Expert System
__ Manager of a department using the Expert System
__ Other User

Performance Criteria

Performance of the Experts

Question Numbers: 10, 49
Total Responses: 35

If human experts currently perform (or previously performed) the task, how often is the expert(s) expected to give the correct answer?

__ Task not performed by human
11 "Correct" defined by expert
10 > 99%
_7 95% to 99%
__ 90% to 95%
_2 80% to 90%
__ 60% to 80%
_1 40% to 60%
_1 Other (100%)
_3 I don't know

Expected Performance of the System

Question Numbers: 11, 50
Total Responses: 34

How often is the Expert System expected to provide the correct answer?

10 100%
_9 > 99%
_4 95% to 99%
_7 90% to 95%
__ 80% to 90%
__ 60% to 80%
__ 40% to 60%
_2 Other
_2 I don't know

Actual Performance of the System

Question Numbers: 12, 51
Total Responses: 32

What is *your estimate* of how often the Expert System **actually** provides the correct answer?

_3 100%
_5 > 99%
_6 95% to 99%
_5 90% to 95%
_5 80% to 90%
_4 60% to 80%
_1 40% to 60%
_1 Other (< 40%)
_2 I don't know

Expected Problem Space Coverage

Question Numbers: 8, 47
Total Responses: 34

How much of the problem space is the Expert System **expected** to cover?

_9 100%
_8 > 99%
_4 95% to 99%
_4 90% to 95%
_3 80% to 90%
_2 60% to 80%
_1 40% to 60%
_1 Other (25%)
_2 I don't know

Actual Problem Space Coverage

Question Numbers: 9, 48
Total Responses: 31

What is *your estimate* of the problem space coverage **actually** provided by the Expert System?

_4 100%
_2 > 99%
_5 95% to 99%
_3 90% to 95%
_8 80% to 90%
_6 60% to 80%
_3 40% to 60%
_2 Other (5%, < 40%)
_3 I don't know

Requirements Definition

Requirements Format

Question Numbers: 13, -
Total Responses: 30

What was the basis for determining how the system was to behave? Please indicate the primary basis with a '*' and check all other applicable basis (if any).

Note: The number of times the choice was selected as primary basis is given in parentheses after the number of times the choice was selected.

_5 (_1) A pre-existing document
10 (_2) A requirements document completed as part of development.
_3 (__) Some other developed document
12 (_3) A prototype of the system
24 (16) Expert consultation
_5 (__) (user feedback, (2) similar tools)

Requirements Difficulty

Question Numbers: 14, -
Total Responses: 29

How difficult was it to develop the original concept of what the system was supposed to do?

_1 Trivial
_6 Easy
15 Medium
_7 Hard
__ Impossible

Availability of the Expert(s)

Question Numbers: 17, -
Total Responses: 26

If the system was not developed by the expert, how much interaction was there between the expert(s) and the development team?

_1 System was developed by expert
_2 Constant
_9 Frequent
_9 Regular
_5 Occasional
__ None

Number of Experts

Question Numbers: 15, -
Total Responses: 30

Was more than one expert consulted during the development of the system?

_6 System was developed by expert
_4 Single expert
_9 Multiple experts with lead
_6 Committee of experts
_5 Other (no experts, experts as available, (2) multiple changing experts)

Agreement Among Experts

Question Numbers: 16, 61
Total Responses: 30

If more than one expert was available for consulting, how often did the experts agree on what results the Expert System was supposed to provide?

_5 A single expert was involved
_5 Always agree
19 Agree 84% of the time (range 50%-99%)

Expert in User Organization

Question Numbers: -, 52
Total Responses: 5

Was the expert(s) a member of the user organization?

_5 Yes
__ No

__ User organization provided some expertise

Developers in User Organization

Question Numbers: 18, 53
Total Responses: 33

Was the developer(s) of the Expert System part of the user organization?

12 Yes
13 No
_8 Some development provided by user organization

Development Information

Development Life-Cycle Used

Question Numbers: 19, -
Total Responses: 30

Please indicate which development model was used for developing the Expert System.

_3 Requirements gathering preceded Design, Implementation, and Test (Traditional waterfall life-cycle).
_4 Requirements gathered before development of a prototype. A second requirements activity preceded Design, Implementation, and Test.
12 Repetition of the Requirements, Design, Rule Generation, and Prototyping phases until production system (final prototype) was developed.
_8 No effort was made to follow a particular model.
_1 Other

Languages and Tools Used

Question Numbers: 20, -
Total Responses: 30

What was the primary language/tool for each part of the Expert System?

Note: The most frequent languages/tools are reported after the choice as: "frequency - language/tool."

26 Knowledge Structures (9 - CLIPS, 7 - LISP, others)
27 Inference Engine (8 - LISP, 8 - CLIPS, 3 - Knowledge Tool, others)
25 Interface Code (12 - C, 7 - LISP, others)

Size of the System

Question Numbers: 22, -
Total Responses: 30

Since Knowledge Bases can be written using several type of Knowledge Structures, please indicate how many of the following structures were used. If another type of structure was used, please describe it and how many were used.

Note: The number of times that a value was given for each choice is provided in parentheses following the number of times that the choice was selected. The range of the responses is given in parentheses after each choice.

25 (14) 184 Rules (range 30-500)
11 (_2) 63 Frames (range 6-120)

11 (_6) 283 Facts (range 100-600)
_7 (_5) 109 Parameters (range 30-312)
_1 (_1) 35K Statements
4 (_0) Other

Total Development Effort

Question Numbers: 29, -
Total Responses: 26

How much effort was expended in developing the system, including evaluation activities performed by the developers? 42 (range 1-300) person/months.

Detailed Development Effort

Question Numbers: 21, -
Total Responses: 29

What percentage of the total development effort was dedicated to each part of the Expert System?

Note: The number of times that a choice was selected is provided in parentheses before the average percentage of effort dedicated to the selected choice. The range of the responses is given in parentheses after each choice.

(29) 54 % Knowledge Structures (range 10%-100%)
(_9) 11 % Inference Engine (range 5%-80%)
(28) 36 % Interface Code (range 10%-80%)

System Sensitivity

Question Numbers: 24, -
Total Responses: 30

When changes were made to the knowledge structures, how often did some unexpected result occur?

_1 Never
20 Occasionally
_7 Frequently
_2 Usually
__ Always

V&V Activities Performed

V&V Activities during development

Question Numbers: 28, -
Total Responses: 30

What testing activities were performed on the executing system? (indicate any that apply)

_1 No evaluation was performed
17 Checked by expert(s)
23 Compared with expected results
12 Structural testing (e.g. cover all rules)
_5 Other

V&V Activities after development

Question Numbers: 33, 60
Total Responses: 26

What testing activities were performed on the executing system **before the system was delivered to the users**? (indicate any that apply)

_1 No evaluation was performed
19 Checked by expert(s)
22 Compared with expected results
15 User acceptance
10 System run in parallel
_3 Other

Development effort was spent on V&V

Question Numbers: 30, -
Total Responses: 16

How much of the development effort was spent on evaluation? 19 % (range 0%-60%)

V&V of Knowledge Structures

Question Numbers: 25, -
Total Responses: 29

What evaluation activities were performed on the Knowledge Structures? (indicate any that apply)

_2 No evaluation was performed
15 Desk checking
_4 Formal inspections
16 Checked by expert(s)
12 Structural testing (e.g. cover all rules)
_7 Other

V&V of Inference Engine

Question Numbers: 26, -
Total Responses: 28

What evaluation activities were performed on the Inference Engine? (indicate any that apply)

19 No evaluation was performed (ES shell was used)
_4 No evaluation was performed
__ Desk checking
_1 Formal inspections
_3 Structural testing
_4 Other

V&V of Interface Code

Question Numbers: 27, -
Total Responses: 28

What evaluation activities were performed on the Interface Code? (indicate any that apply)

_5 No evaluation was performed

14 Desk checking
_5 Formal inspections
14 Structural testing (branch or path)
_8 Other

Difficulty of V&V

Question Numbers: 35, 62
Total Responses: 27

Compared to conventional software testing efforts, how difficult was the evaluation of the Expert System?

__ Trivial
_5 Easy
10 Medium
12 Hard
__ Impossible
__ No evaluation was done

Separate V&V group

Question Numbers: 31, -
Total Responses: 26

Did a separate organization evaluate the Expert System before it was delivered to the users?

_5 Yes, there was a separate evaluation organization.
21 No, there was not a separate evaluation organization.

Independent V&V Effort

Question Numbers: 32, 59
Total Responses: 5

If there was a separate evaluation team, how much effort was expended by the team in evaluating the correctness of the Expert System?

(2) 1.7 (range .5-3) person/months reported by developers
(3) 16 (range (5-24) person/months reported by users

Operational or Prototype System

Question Numbers: 3, 43
Total Responses: 35

Is the Expert System operational or is it a prototype?

15 Operational system
19 Prototype system
_1 Operational prototype (write in)

System Criticality

Question Numbers: 37, 55
Total Responses: 34

How reliable is the Expert System required to be?

_4 Trusted with human life
_8 Trusted with mission objectives
17 As reliable as the expert

10 Assists the expert
_8 Assists the user
__ Other

V&V Issues Encountered

Known Issues Actually Encountered

Question Numbers: 36, 63
Total Responses: 32

Many people feel that some development issues are more of a problem with Expert Systems than with conventional systems. Which (if any) of the following were problems during implementation or test of this Expert System?

_8 Understandability and readability of knowledge structures
19 Determining test coverage for knowledge structures
11 Modularity/Design of knowledge structures
21 Knowledge validation
_2 Analysis of Certainty Factors
_6 Validating the inference engine
13 Real-time performance analysis
16 Complexity of the Problem
_8 Certification
_7 Configuration Management
_3 Other

Certainty Factors

Question Numbers: 7, -
Total Responses: 30

Does the Expert System include certainty factors?

_2 Yes
26 No
_2 I don't know

Configuration Management

Question Numbers: 34, -
Total Responses: 16

How were changes to the Expert System distributed to the users?

_3 User updated system at developer's direction
_7 Developers made changes to users' system
_1 Untested system distributed to users
_4 Tested system distributed to the users
_1 Configuration management group distributes system
__ Other

Expertise Implementation Difficulty

Question Numbers: 23, -
Total Responses: 30

Aside from any difficulties in developing the original concept, how difficult was it to express the behavior (through the Knowledge Structures) of the expert?

__ Trivial
_3 Easy
17 Medium
10 Hard
__ Impossible

Expected System Use

Question Numbers: 39, 57
Total Responses: 26

How many people are **expected** to make use of the Expert System? 279 (range 3-2000)

Actual System Use

Question Numbers: 40, 58
Total Responses: 12

How frequently are the (expected) users **actually** using the system? (Numbers may add up to more than 100% if the actual number of users is greater than the expected users.)

Note: The number of times a value was given is provided in parentheses before the percentage of use corresponding to each choice.

(_4) 9 % use the system more than expected (range 5%-60%)
(11) 46 % use the system about as much as expected (range 10%-80%)
(11) 23 % use the system less than expected (range 10%-90%)
(_7) 22 % do not use the system (range 10%-90%)

Perceived System Reliability

Question Numbers: 38, 56
Total Responses: 35

Does the Expert System seem to be more reliable or less reliable than conventional systems that are in use?

_1 Significantly more reliable
_9 More reliable
__ Slightly more reliable
_6 Similar reliability
_1 Slightly less reliable
_1 Less reliable
__ Significantly less reliable
12 No comparison is available
_5 I don't know

User Trust

Question Numbers: -, 54
Total Responses: 5

Why do you believe the results that the system gives?

_1 Expert says it is correct
_3 Participated in evaluation
__ Someone I trust did evaluation
_5 Personal use and checking
_1 User acceptance
__ I don't trust the results
__ Other

Recommendations

The recommendations from the survey results are separated into two categories:

Direct Recommendations

Recommendations in this category are directly supported by the survey results. These recommendations include:

- Develop Requirements for Expert System Verification and Validation
- Address Most Often Encountered Issues
- Recommend a Life Cycle for Expert Systems Development

Inferred Recommendations

Recommendations in this category can be inferred from the survey results by analyzing relationships among the responses. These recommendations include:

- Address Readability and Modularity Issues
- Address Configuration Management Issue
- Develop Criteria to Classify Expert Systems by Intended Use
- Investigate Applicability of Analysis Tools

Following each general recommendation is an explanation of what was observed in the survey results. After this explanation is a list of specific recommendations which address all the observations. Each specific recommendation in the "Direct Recommendations" section is followed by a list of supporting phrases from "Summary of Results" on page 7.

Direct Recommendations

Develop Requirements for Expert System Verification and Validation

The major goal of this survey task was to discover and document the current state of the practice in Verification and Validation of Expert Systems. Based on the survey results, it appears that much can be done to improve the practice. The lack of requirements for performing V&V on ESs was manifested in several forms:

- The V&V activities performed were very inconsistent, ranging from none to very many, and the sets of activities performed were very diverse.
- The reliance on expert consultation as the only source of requirements was extremely high.
- The reliance on experts to perform V&V activities on the knowledge base, interface code, and executing systems was very high.
- The low expected and actual performance levels for many of the expert systems was surprising. It is unlikely that conventional software systems that exhibited this level of performance would gain wide acceptance. (For example, many reported that the ES provides the correct answer less than 90 % of the time. Most conventional software reliability is rated as a series of '9's, e.g., 4 '9's means the correct answer is given > 99.99 % of the time.)
- In those cases where the expected behavior of the system was not strictly defined by expert consultation, a large number of systems relied on prototypes.

This is significant because prototype systems receive less V&V than operational systems, but are then used to define the behavior of operational systems.

Each of the above observations can be directly attributed to three factors:

1. There is a general lack of understanding on how to V&V ESs. Generally, it is not known what V&V activities are to be performed, when the activities should be performed, or how the activities can be accomplished.
2. There is little understanding of how requirements for an ES should be generated and documented. It could be argued that this is a development issue, but without documented expected behavior, there is no possibility of performing adequate V&V.
3. A large number of expert systems are prototypes for which V&V receives little consideration.

Recommendations

1. Develop recommendations and/or guidelines for Verification and Validation of Expert Systems. (Since such a significant amount of research has been devoted to V&V of traditional software, it may be appropriate to approach this task as a set of modifications to current conventional software V&V requirements.)

    "Of thirty respondents, twenty-four indicated that expert consultation was a basis for determining the behavior of the system."

    "Most V&V activities relied on comparison with expected results and expert checking"

    "In most cases, there was not a separate group to perform V&V"

2. Initial efforts to define V&V requirements should be focused on diagnostic systems, since a large majority of the systems surveyed performed diagnostic services.

    "Most ... perform Diagnosis (82%) ..."

3. Research the process of converting prototype ESs into operational systems. A large number of respondents indicated that they were either building prototypes for later conversion into operational systems, or building operational systems based on prototypes.

    "Of thirty respondents ... Fourteen respondents indicated that prototypes or similar tools were used for the requirements"

    "Fifty-three percent of the respondents indicated that the ES was a prototype system."

Address Most Often Encountered Issues

All of the known issues with performing V&V on Expert Systems were cited at least once in the survey. A small group of issues, however, were cited significantly more often than others and included:

1. Knowledge validation,
2. Determining test coverage, and
3. Complexity of the problem

The first two issues are well understood and are active research areas. These research areas should be matured so that they solutions to these issues can be provided.

The complexity issue is not as well understood. These is considerable opinion that the types of problems addressed by ESs are significantly harder than the problems addressed by conventional software. Others maintain the apparent difficulty is attributed to the lack of requirements (see above). In either case, there does not seem to be a way to approach the complexity issue without considering it in the context of the readability and modularity issues, as done in "Address Readability and Modularity Issues" on page 23.

Recommendations

1. Develop methods and/or tools to support the knowledge validation activity.

   "The known issues most often cited as problems were: knowledge validation (66%) ..."

2. Develop tools and/or methods to support the determination of test coverage.

   "The known issues most often cited as problems were: ... test coverage determination (59%) ..."

Recommend a Life Cycle for Expert Systems Development

The most common Life Cycle applied to the development of the ESs included in this survey was the Cyclic model. In the Cyclic model, the stages of requirements, design, knowledge base development, and test are repeated until the final system is developed. The testing activities at the end of each cycle (except the last) lead to the refinement of the requirements that will be used in the successive cycle. Several variations, including some with a fixed number of cycles, have been proposed.

A large number of respondents, however, indicated that no attempt was made to follow any model. If no model is being followed, there is little opportunity to apply V&V activities at the appropriate points during development. Clearly, any life cycle guidelines would be of benefit in these situations. Multiple life-cycle approaches, or a single very flexible life-cycle should be recommended.

Recommendation

1. Multiple life cycle models, or a single, very flexible life cycle model should be recommended for development of ESs. (The high incidence of prototypes leading to operational systems suggests that the cyclic model should be recommended. Rapid prototyping could be treated as a special case of the cyclic model.)

   "The most frequent (40%) Life-Cycle model used is the Cyclic Model ... however, 27% ... stated that no model was followed."

   "Of thirty respondents ... Fourteen respondents indicated that prototypes or similar tools were used for the requirements"

   "Fifty-three percent of the respondents indicated that the ES was a prototype system."

Inferred Recommendations

Address Readability and Modularity Issues

Readability and modularity were expected to be significant issues, but were not the most frequently cited problems. Further analysis of the survey results indicate that the readability and modularity issues may have been reported as other problems. This analysis includes the following observations:

- As often as not, people chose modularity or readability as problems, but not both. This seems to indicate that many respondents do not see the relationship between the two.
- Similarly, as often as not, people picked test coverage determination without picking modularity, so the apparent relationship between there two issues was not established.
- The lack of reported relationships between the readability, modularity, and test coverage issues is very confusing, implying, for instance, that a rule can be understood but a test scenario for it can not be developed.
- Readability and complexity of the problem were very rarely chosen together. That is, the developer recognizes that the ES was complicated but attributed this complexity either to the problem or to the solution, but not both. It is questionable that the complexity of the problem and the complexity of the solution can be easily distinguished. (The emergence of Object-oriented programming languages is due, in part, to the claim that conventional languages cause programming complexities which are erroneously attributed to problem complexity.)

If the number of times each of these issues were reported are added together, the collection of issues becomes a very frequently cited problem. Since these issues are so closely interrelated, they should be addressed as a single issue. Therefore, the problem of reducing overall complexity (problem/solution) is a very important issue.

Recommendation

1. Develop methods and/or tools to support the readability, modularity, and problem complexity issue.

Address Configuration Management Issue

Configuration management was an infrequently cited problem. However, the survey results also show that in practice the applied CM, while sometimes quite good, was generally poor (changes to the knowledge base were not well managed). This contradiction is probably due to the high frequency of prototypes and "in development" responses to the survey. While there are certain applications for which CM may never be a significant issue, certainly there are applications for which CM is a very important issue.

Recommendation

1. Identify the differences between CM of conventional software systems and CM of expert systems. It is not immediately obvious that there are differences.

Develop Criteria to Classify Expert Systems by Intended Use

The survey results indicate that there is a very diverse set of applications which are utilizing ES technology. At least the following types of applications exist:

Expert Clone
: Provides expert assistance to a human user. The expert is usually available if the ES does not provide the correct results. The major uses of this type of include: education and capture of true institutional knowledge.

Expert Assistant
: Allows the user, typically an expert, to concentrate on the more important aspects of the task. These ESs typically serve as filtering mechanisms.

Autonomous
: Limited supervision is applied to the ES. In additional to providing filtering, these systems typically develop and execute plans to handle situations.

A subcategory of Autonomous ESs are time critical ESs. These ESs exist primarily because experts can not interpret data efficiently enough to perform the task in the allotted time.

Self-modifying autonomous
: Part of the planned execution is to modify its knowledge base to respond to certain situational data. The application of V&V to this type of problem is currently uncertain.

Traditional Software Problem
: Some conventional problems (e.g. discrete event simulation). are more conveniently implemented using expert system shells

It is apparent that because of this diversity, a single set of V&V requirements is probably undesirable. Development of classification criteria allows a simplification of ES V&V requirements. In addition to simplification, classification allows the development of requirements to be concentrated on the types of applications of interest.

Recommendations

1. Develop classification criteria to distinguish among expert systems which require different V&V approaches.
2. Concentrate initial V&V requirements definition effort on autonomous systems, since these systems are likely the most critical.

Investigate Applicability of Analysis Tools

A very large number of respondents indicated that experts were the primary source of requirements and verification. Several of the previous recommendations would reduce this dependence, but there is a class of expert system applications for which expert consultation will continue to be the leading source.

Recommendations

1. Determine if a there is a communication problem between the experts and the knowledge engineers / expert system developers.
2. If a communication problem exists, investigate the applicability of Knowledge Base to natural language translators as a possible solution.

Appendix A. Expert Systems Evaluation Questionnaire (Developer)

By filling out this NASA funded questionnaire, you can help define the state-of-the-practice in the formal evaluation of Expert Systems on current NASA and industry applications. The information that you provide will be merged with the information from all other surveyed projects for the purpose of recommending future research and development activities. Individual responses are used solely as input to this information merging process. Each survey participant will be sent a copy of the final survey results.

Expert System applications are becoming more prevalent in fields where proper functioning is essential, such as the aerospace, medical, and financial industries. It is widely claimed that Expert Systems are not as rigorously evaluated as traditional software because of unique, unresolved evaluation issues. To ensure the continued and safe deployment of Expert Systems into critical areas, adequate evaluation techniques which address these issues must be developed and performed.

Instructions

The following questions concern your experiences with an Expert System, either as a developer or as the manager of the development effort. Feel free to indicate your answers in any way you like. Some of the choices on the multiple choice questions have places to fill in additional information; please indicate the choice and include the additional information, if possible. If you have any comments about the questions or your answers, please write them in the left margin.

Analysis of the responses may indicate that further discussion is required for complete understanding of the issues encountered during the evaluation process. Discussions will be held either as short one-on-one meetings or by telephone. Would you be available, at your convenience, to discuss the evaluation process in more detail?

Yes I am available for discussions.

Name ______________________________

Phone _______________

No I am not available for discussions.

If you have any questions regarding this questionnaire, please contact Keith Kelley at (713) 282-7303. If possible, please return completed questionnaires within one week of receipt to:

Keith Kelley
MC 6606
IBM Federal Sector Division
3700 Bay Area Blvd.
Houston, Tx. 77058-1199

Definitions

Certainty factors

Some problems require the use of certainty factors (also called probabilities, or fuzzy logic) in their processing. Facts which contain certainty factors have the form: "if a is true, then there is an x% chance that b is true."

Expert

The person who provides the knowledge that is to be captured in the Expert System.

Inference engine

Processes the knowledge structures to infer a set of output facts from a set of input facts. Examples of commercial systems are CLIPS and ESE.

Interface code

Used to supplement the inference process. Examples are interfacing the inference engine to a device, and performing arithmetic calculations.

Knowledge structures

Declarative part of the Expert System which represents the knowledge (typically called the Knowledge Base). Examples are frames and rules.

Problem space

The total number of cases which could potentially be addressed by the Expert System.

Problem space coverage

The percentage of the problem space that is addressed by the Expert System. For example, if the Expert System is supposed to be able to diagnose 100 malfunctions, but the total number of malfunctions is known to be 200, the problem space coverage is 50%.

Questions

1. What is the name of the Expert System you were/are involved with?

   ____

2. Were you a developer of the Expert System or the manager of the development organization?

   a. Developer of Expert System
   b. Manager of Expert System development organization
   c. Other ______________________________

3. Is the Expert System operational or is it a prototype?

   a. Operational system b. Prototype system

4. Briefly describe what the expert system does.

   ____

   ____

   ____

5. What field does the problem belong to?

a.	Aerospace	g.	Medical
b.	Financial	h.	Personnel
c.	Information Systems	i.	Research
d.	Hardware	j.	Service
e.	Manufacturing	k.	Software
f.	Marketing	l.	Other ________________

6. Which of the following items best describes the kind of problem the Expert System addresses? Please indicate primary purpose with a '*' and check all other applicable purposes (if any).

   a. Design - Configuring objects under constraints
   b. Repair - Executing plans to administer prescribed remedies
   c. Control - Governing overall system behavior
   d. Planning - Designing actions
   e. Diagnosis - Inferring system malfunctions from observables
   f. Debugging - Prescribing remedies for malfunctions
   g. Prediction - Inferring likely consequences of given situations
   h. Monitoring - Comparing observations to expected outcomes
   i. Instruction - Diagnosing, debugging, and repairing behavior
   j. Interpretation - Inferring situation descriptions from sensor
   k. Classification - Categorizing objects by properties data

7. Does the Expert System include certainty factors?

a.	Yes	c.	I don't know
b.	No		

8. How much of the problem space is the Expert System **expected** to cover?

a.	100%	f.	60% to 80%
b.	> 99%	g.	40% to 60%
c.	95% to 99%	h.	Other ________%
d.	90% to 95%	i.	I don't know
e.	80% to 90%		

9. What is *your estimate* of the problem space coverage **actually** provided by the Expert System?

a.	Same as expected	f.	80% to 90%
b.	100%	g.	60% to 80%
c.	> 99%	h.	40% to 60%
d.	95% to 99%	i.	Other ________%
e.	90% to 95%	j.	I don't know

Questions 10 through 12 are concerned with the percentage of problems within the problem space (covered by the Expert System) that are answered correctly.

10. If human experts currently perform (or previously performed) the task, how often is the expert(s) expected to give the correct answer?

    a. Task not performed by human
    b. "Correct" defined by expert
    c. > 99%
    d. 95% to 99%
    e. 90% to 95%
    f. 80% to 90%
    g. 60% to 80%
    h. 40% to 60%
    i. Other ________%
    j. I don't know

11. How often is the Expert System expected to provide the correct answer?

    a. 100%
    b. > 99%
    c. 95% to 99%
    d. 90% to 95%
    e. 80% to 90%
    f. 60% to 80%
    g. 40% to 60%
    h. Other ________%
    i. I don't know

12. What is *your estimate* of how often the Expert System **actually** provides the correct answer?

    a. 100%
    b. > 99%
    c. 95% to 99%
    d. 90% to 95%
    e. 80% to 90%
    f. 60% to 80%
    g. 40% to 60%
    h. Other ________%
    i. I don't know

13. What was the basis for determining how the system was to behave? Please indicate the primary basis with a '*' and check all other applicable basis (if any).

    a. A pre-existing document ____________________
    b. A requirements document completed as part of development.
    c. Some other developed document ____________________
    d. A prototype of the system
    e. Expert consultation
    f. Other ____________________

14. How difficult was it to develop the original concept of what the system was supposed to do?

    a. Trivial
    b. Easy
    c. Medium
    d. Hard
    e. Impossible

15. Was more than one expert consulted during the development of the system?

    a. System was developed by expert
    b. Single expert
    c. Multiple experts with lead
    d. Committee of experts
    e. Other ____________________

16. If more than one expert was available for consulting, how often did the experts agree on what results the Expert System was supposed to provide?
    a. A single expert was involved
    b. Always agree
    c. Agree ______% of the time.

17. If the system was not developed by the expert, how much interaction was there between the expert(s) and the development team?
    a. System was developed by expert
    b. Constant
    c. Frequent
    d. Regular
    e. Occasional
    f. None

18. Was the developer(s) part of the user organization?
    a. Yes
    b. No
    c. Some developers were in the user organization

19. Please indicate which development model was used for developing the Expert System.
    a. Requirements gathering preceded Design, Implementation, and Test (Traditional waterfall life-cycle).
    b. Requirements gathered before development of a prototype. A second requirements activity preceded Design, Implementation, and Test.
    c. Repetition of the Requirements, Design, Rule Generation, and Prototyping phases until production system (final prototype) was developed.
    d. No effort was made to follow a particular model.
    e. Other ____

20. What was the primary language/tool for each part of the Expert System?
    a. Knowledge Structures ________________________________
    b. Inference Engine ________________________________
    c. Interface Code ________________________________

21. What percentage of the total development effort was dedicated to each part of the Expert System?
    a. Knowledge Structures _________%
    b. Inference Engine _________% (If an Expert System Shell was used, this value should be 0%.)
    c. Interface Code _________%

22. Since Knowledge Bases can be written using several type of Knowledge Structures, please indicate how many of the following structures were used. If another type of structure was used, please describe it and how many were used.

    a. Rules ________
    b. Frames ________
    c. Facts ________
    d. Parameters ________
    e. Statements ________
    f. Other (#)________ of ________

23. Aside from any difficulties in developing the original concept, how difficult was it to express the behavior (through the Knowledge Structures) of the expert?

    a. Trivial
    b. Easy
    c. Medium
    d. Hard
    e. Impossible

24. When changes were made to the knowledge structures, how often did some unexpected result occur?

    a. Never
    b. Occasionally
    c. Frequently
    d. Usually
    e. Always

Questions 25 through 28 are concerned with the evaluation activities performed during development.

25. What evaluation activities were performed on the knowledge Structures? (indicate any that apply)

    a. No evaluation was performed
    b. Desk checking
    c. Formal inspections
    d. Checked by expert(s)
    e. Structural testing (e.g. cover all rules)
    f. Other ________________

26. What evaluation activities were performed on the Inference Engine? (indicate any that apply)

    a. No evaluation was performed
    b. Desk checking
    c. Formal inspections
    d. Structural testing
    e. Other ________________

27. What evaluation activities were performed on the Interface Code? (indicate any that apply)

    a. No evaluation was performed
    b. Desk checking
    c. Formal inspections
    d. Structural testing (branch or path)
    e. Other ________________

28. What testing activities were performed on the executing system? (indicate any that apply)

    a. No evaluation was performed
    b. Checked by expert(s)
    c. Compared with expected results
    d. Structural testing (e.g. cover all rules)
    e. Other ________________

29. How much effort was expended in developing the system, including evaluation activities performed by the developers? ________ person/months.

30. How much of the development effort was spent on evaluation? ________ %.

31. Did a separate organization evaluate the Expert System before it was delivered to the users?

    a. Yes, there was a separate evaluation organization.
    b. No, there was not a separate evaluation organization.

32. If there was a separate evaluation team, how much effort was expended by the team in evaluating the correctness of the Expert System? ________ person/months.

33. What testing activities were performed on the executing system **before the system was delivered to the users**? (indicate any that apply)

    a. No evaluation was performed
    b. Checked by expert(s)
    c. Compared with expected results
    d. User acceptance
    e. System run in parallel
    f. Other ________________

34. How were changes to the Expert System distributed to the users?

    a. User updated system at developer's direction
    b. Developers made changes to users' system
    c. Untested system distributed to users
    d. Tested system distributed to the users
    e. Configuration management group distributes system
    f. Other ____

35. Compared to conventional software testing efforts, how difficult was the evaluation of the Expert System?

    a. Trivial
    b. Easy
    c. Medium
    d. Hard
    e. Impossible
    f. No evaluation was done

36. Many people feel that some development issues are more of a problem with Expert Systems than with conventional systems. Which (if any) of the following were problems during implementation or test of this Expert System?

    a. Understandability and readability of knowledge structures
    b. Determining test coverage for knowledge structures
    c. Modularity/Design of knowledge structures
    d. Knowledge validation
    e. Analysis of Certainty Factors
    f. Validating the inference engine
    g. Real-time performance analysis
    h. Complexity of the Problem
    i. Certification
    j. Configuration Management
    k. Other ______________________________

37. How reliable is the Expert System required to be?

    a. Trusted with human life
    b. Trusted with mission objectives
    c. As reliable as the expert
    d. Assists the expert
    e. Assists the user
    f. Other ____________

38. Does the Expert System seem to be more reliable or less reliable than conventional systems that are in use?

    a. Significantly more reliable
    b. More reliable
    c. Slightly more reliable
    d. Similar reliability
    e. Slightly less reliable
    f. Less reliable
    g. Significantly less reliable
    h. No comparison is available
    i. I don't know

39. How many people are **expected** to make use of the Expert System?

    ____

40. How frequently are the (expected) users **actually** using the system? (Numbers may add up to more than 100% if the actual number of users is greater than the expected users.)

    a. ________% use the system more than expected

    b. ________% use the system about as much as expected

    c. ________% use the system less than expected

    d. ________% do not use the system

Appendix B. Expert Systems Evaluation Questionnaire (User)

By filling out this NASA funded questionnaire, you can help define the state-of-the-practice in the formal evaluation of Expert Systems on current NASA and industry applications. The information that you provide will be merged with the information from all other surveyed projects for the purpose of recommending future research and development activities. Individual responses are used solely as input to this information merging process. Each survey participant will be sent a copy of the final survey results.

Expert System applications are becoming more prevalent in fields where proper functioning is essential, such as the aerospace, medical, and financial industries. It is widely claimed that Expert Systems are not as rigorously evaluated as traditional software because of unique, unresolved evaluation issues. To ensure the continued and safe deployment of Expert Systems into critical areas, adequate evaluation techniques which address these issues must be developed and performed.

Instructions

The following questions concern your experiences with an Expert System, either as a user or as the manager of a department that uses Expert System. Feel free to indicate your answers in any way you like. Some of the choices on the multiple choice questions have places to fill in additional information; please indicate the choice and include the additional information, if possible. If you have any comments about the questions or your answers, please write them in the left margin.

Analysis of the responses may indicate that further discussion is required for complete understanding of the issues encountered during the evaluation process. Discussions will be held either as short one-on-one meetings or by telephone. Would you be available, at your convenience, to discuss the evaluation process in more detail?

Yes I am available for discussions.

Name ______________________________

Phone ______________

No I am not available for discussions.

If you have any questions regarding this questionnaire, please contact Keith Kelley at (713) 282-7303. If possible, please return completed questionnaires within one week of receipt to:

Keith Kelley
MC 6606
IBM Federal Sector Division
3700 Bay Area Blvd.
Houston, Tx. 77058-1199

Definitions

Expert
The person who provides the knowledge that is to be captured in the Expert System.

Inference engine
Processes the knowledge structures to infer a set of output facts from a set of input facts. Examples of commercial systems are CLIPS and ESE.

Knowledge structures
Declarative part of the Expert System which represents the knowledge (typically called the Knowledge Base). Examples are frames and rules.

Problem space
The total number of cases which could potentially be addressed by the Expert System.

Problem space coverage
The percentage of the problem space that is addressed by the Expert System. For example, if the Expert System is supposed to be able to diagnose 100 malfunctions, but the total number of malfunctions is known to be 200, the problem space coverage is 50%.

Questions

41. What is the name of the Expert System you were/are involved with?

    ____

42. Are you a user of the Expert System or the manager of a department which uses the Expert System?

    a. User of the Expert System
    b. Manager of a department using the Expert System
    c. Other ______________________________

43. Is the Expert System operational or is it a prototype?

    a. Operational system
    b. Prototype system

44. Briefly describe what the expert system does.

    ____

    ____

    ____

45. What field does the problem belong to?

    a. Aerospace
    b. Financial
    c. Information Systems
    d. Hardware
    e. Manufacturing
    f. Marketing
    g. Medical
    h. Personnel
    i. Research
    j. Service
    k. Software
    l. Other ____________________

46. Which of the following items best describes the kind of problem the Expert System addresses? Please indicate primary purpose with a '*' and check all other applicable purposes (if any).

    a. Design - Configuring objects under constraints
    b. Repair - Executing plans to administer prescribed remedies
    c. Control - Governing overall system behavior
    d. Planning - Designing actions
    e. Diagnosis - Inferring system malfunctions from observables
    f. Debugging - Prescribing remedies for malfunctions
    g. Prediction - Inferring likely consequences of given situations
    h. Monitoring - Comparing observations to expected outcomes
    i. Instruction - Diagnosing, debugging, and repairing behavior
    j. Interpretation - Inferring situation descriptions from sensor data
    k. Classification - Categorizing objects by properties

47. How much of the problem space is the Expert System **expected** to cover?

    a. 100%
    b. > 99%
    c. 95% to 99%
    d. 90% to 95%
    e. 80% to 90%
    f. 60% to 80%
    g. 40% to 60%
    h. Other _________%
    i. I don't know

48. What is *your estimate* of the problem space coverage **actually** provided by the Expert System?

    a. Same as expected
    b. 100%
    c. > 99%
    d. 95% to 99%
    e. 90% to 95%
    f. 80% to 90%
    g. 60% to 80%
    h. 40% to 60%
    i. Other _________%
    j. I don't know

Questions 49 through 51 are concerned with the percentage of problems within the problem space (covered by the Expert System) that are answered correctly.

49. If human experts currently perform (or previously performed) the task, how often is the expert(s) expected to give the correct answer?

    a. Task not performed by human
    b. "Correct" defined by expert
    c. > 99%
    d. 95% to 99%
    e. 90% to 95%
    f. 80% to 90%
    g. 60% to 80%
    h. 40% to 60%
    i. Other _________%
    j. I don't know

50. How often is the Expert System expected to provide the correct answer?

    a. 100%
    b. > 99%
    c. 95% to 99%
    d. 90% to 95%
    e. 80% to 90%
    f. 60% to 80%
    g. 40% to 60%
    h. Other _________%
    i. I don't know

51. What is *your estimate* of how often the Expert System **actually** provides the correct answer?

a. 100%
b. > 99%
c. 95% to 99%
d. 90% to 95%
e. 80% to 90%
f. 60% to 80%
g. 40% to 60%
h. Other ________%
i. I don't know

52. Was the expert(s) a member of the user organization?

a. Yes
b. No
c. User organization provided some expertise

53. Was the developer(s) of the Expert System part of the user organization?

a. Yes
b. No
c. Some development provided by user organization

54. Why do you believe the results that the system gives?

a. Expert says it is correct
b. Participated in evaluation
c. Someone I trust did evaluation
d. Personal use and checking
e. User acceptance
f. I don't trust the results
g. Other ________________

55. How reliable is the Expert System required to be?

a. Trusted with human life
b. Trusted with mission objectives
c. As reliable as the expert
d. Assists the expert
e. Assists the user
f. Other ________________

56. Does the Expert System seem to be more reliable or less reliable than conventional systems that are in use?

a. Significantly more reliable
b. More reliable
c. Slightly more reliable
d. Similar reliability
e. Slightly less reliable
f. Less reliable
g. Significantly less reliable
h. No comparison is available
i. I don't know

57. How many people are expected to make use of the Expert System?

____

58. How frequently are the (expected) users **actually** using the system? (Numbers may add up to more than 100% if the actual number of users is greater than the expected users.)

    a. _________% use the system more than expected

    b. _________% use the system about as much as expected

    c. _________% use the system less than expected

    d. _________% do not use the system

If you were not involved with evaluating the Expert System, please leave the remaining questions unanswered.

59. How much effort was expended by the evaluation team in evaluating the correctness of the Expert System? _________ person/months.

60. What testing activities were performed on the executing system **before the system was delivered to the users**? (indicate any that apply)

    a. No evaluation was performed
    b. Checked by expert(s)
    c. Compared with expected results
    d. User acceptance
    e. System run in parallel
    f. Other ________________

61. If more than one expert was available for consulting, how often did the experts agree on what results the Expert System is supposed to provide?

    a. No expert was involved
    b. A single expert was involved
    c. Always agree
    d. Agree _____% of the time.

62. Compared to conventional software testing efforts, how difficult was the evaluation of the Expert System?

    a. Trivial
    b. Easy
    c. Medium
    d. Hard
    e. Impossible

63. Many people feel that some development issues are more of a problem with Expert Systems than with conventional systems. Which (if any) of the following were problems during testing of the Expert System?

    a. Understandability and readability of knowledge structures
    b. Determining test coverage for knowledge structures
    c. Modularity/Design of knowledge structures
    d. Knowledge validation
    e. Analysis of Certainty Factors
    f. Validating the inference engines
    g. Real-time performance analysis
    h. Complexity of the Problem
    i. Certification
    j. Other ________________________________